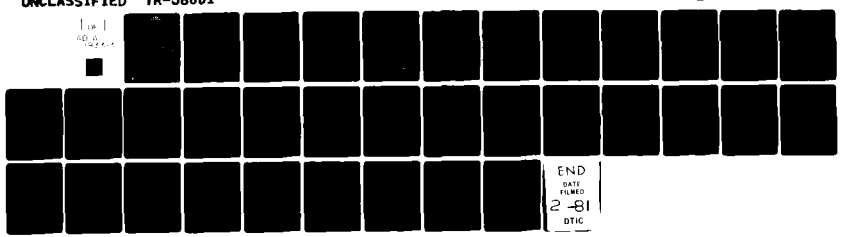
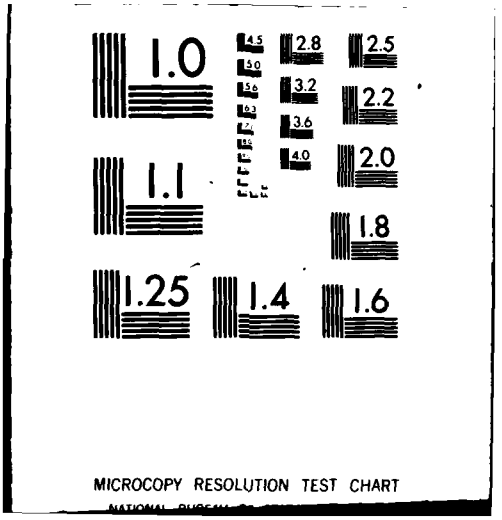
AD-A093 313 MASSACHUSETTS UNIV AMHERST DEPT OF MATHEMATICS AND S--ETC F/G 12/1
OPTIMALITY MEASURES FOR MONOTONE EQUIVARIANT CLUSTER TECHNIQUES--ETC(U)
SEP 80 M JANOWITZ N00014-79-C-0629
UNCLASSIFIED TR-J8001 NL



END
DATE FILMED
2-81
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART

LEVEL II

12

Department of Mathematics & Statistics

University of Massachusetts

AD A093313

# Optimality Measures for Monotone Equivariant Cluster Techniques

by

M. F. Janowitz



Technical Report J8001

September, 1980

# Optimality Measures for Monotone Equivariant Cluster Techniques*

by

M. F. Janowitz**

1. Introduction. This paper is devoted to the investigation of various measures of optimality for cluster techniques. Before proceeding, it is appropriate to discuss the nature of the cluster methods that will be considered.

The starting point is a finite (nonempty) set $P$ equipped with a dissimilarity measure (henceforth known as a DC) $d$. This is merely a mapping $d:P \times P \to R^+$, the set of nonnegative real numbers, such that

(DC1) $d(a,b) = d(b,a)$,

(DC2) $d(a,a) = 0$

for all $a,b \in P$. In the Jardine-Sibson model for cluster analysis [16], a cluster method converts this input DC into a numerically stratified clustering (NSC); this is a mapping $C:R^+ \to \Sigma$, where $\Sigma$ denotes the set of reflexive symmetric relations on $P$, ordered by inclusion, such that

(NSC1) $C$ is isotone,

(NSC2) $C(h) = P \times P$ for some $h \in R^+$,

(NSC3) Corresponding to each element $h$ of $R^+$, there is a $\delta > 0$ such that $C(h) = C(h + \delta)$.

For each $d \in C(P)$, the set of DC's on $P$, there corresponds an NSC $Td$ given by

$$Td(h) = \{(a,b):d(a,b) \leq h\},$$

---

*Presented in part to the Classification Society April 9, 1979.

**Research supported in part by ONR Contract N00014-79-C-0629 as well as by grants from the University of Massachusetts Computer Center.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

12 34

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

14 TR-J8[illegible]1

2. GOVT ACCESSION NO. AD-A093318

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

6 Optimality Measures for Monotone Equivariant Cluster Techniques.

5. TYPE OF REPORT & PERIOD COVERED

9 Technical rept.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

10 M.F. Janowitz

8. CONTRACT OR GRANT NUMBER(s)

15 N00014-79-C-0629

9. PERFORMING ORGANIZATION NAME AND ADDRESS

University of Massachusetts
Amherst, MA 01003

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

121405

11. CONTROLLING OFFICE NAME AND ADDRESS

Procuring Contract Officer
Office of Naval Research
Arlington, VA 22217

11 12. REPORT DATE

Sep[illegible]80

13. NUMBER OF PAGES

33

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

Office of Naval Research Resident Representative, Harvard University
Gordon McKay Laboratory, Room 113
Cambridge, MA 02138

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Cluster analysis, Optimality measure, Cophenetic correlation, Monotone equivariant

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

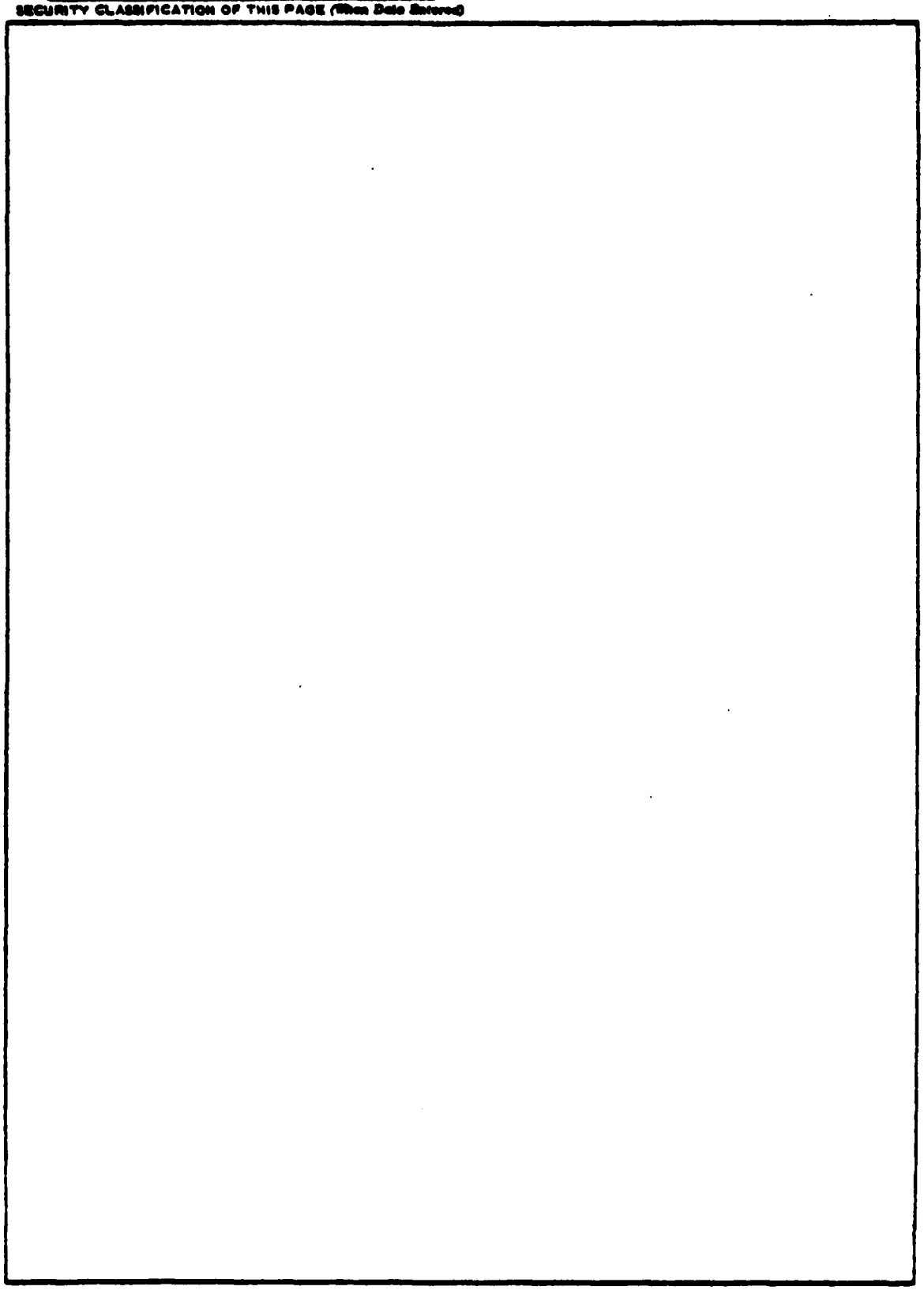
Several commonly used optimality measures are compared and classified as to their performance with monotone equivariant cluster techniques. A new measure is introduced and compared to the earlier measures. The measures are then applied to determining the optimality of several cluster techniques when they are applied to some concrete data that arises from the psychological and social sciences.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

411377 Dm



and Jardine and Sibson argue in [16] that $d \to Td$ is a bijection between $C(P)$ and the set of all NSC's on $P$. For that reason, a cluster method may either be regarded as a mapping $F:C(P) \to C(P)$ or as a mapping $\overline{F}$ on the set of NSC's; here $\overline{F}(Td) = T(Fd)$ for all $d \in C(P)$.

So that the abstract theory of partially ordered sets could be brought to bear on the theory of cluster analysis, an order theoretic model for the subject was introduced in [10] and further developed in [11], [12], [13] and [14]. The idea is to notice that an NSC is the same thing as a residual mapping of $R^+$ into $\Sigma$. Here a mapping $f$ of a partially ordered set $A$ into a partially ordered set $B$ is said to be <u>residual</u> if $f$ is isotone, and there is an isotone mapping $f^*:B \to A$ such that $f^*f(a) \leq a$ and $ff^*(b) \geq b$ for all $a \in A$, $b \in B$. This associated mapping $f^*$ is the <u>residuated</u> mapping determined by $f$. A full treatment of residuated and residual mappings may be found in [1]. Letting $Res(B,A)$ denote the set of residuated mappings of $B$ into $A$, and $Res^+(A,B)$ the residual maps of $A$ into $B$, a cluster method may now be defined as follows. The nonnegative reals are replaced by an abstract join semilattice $L$ with $0$, and the set $\Sigma$ of reflexive symmetric relations by a pair $M$, $N$ of finite partially ordered sets, each of which has a largest element that is not its only element. A cluster method can now be regarded as a function $F:Res(M,L) \to Res(N,L)$ or as a function $\overline{F}:Res^+(L,M) \to Res^+(L,N)$. For purposes of the present work, this level of abstraction is not needed, and we shall view a cluster method as a function on $C(P)$ or as a function on $Res^+(R^+,\Sigma)$.

If one assumes that the input DC $d$ has only ordinal significance, it is shown in [12] that one should use a <u>monotone equivariant</u> cluster method. This is a cluster method $F$ having the property that for each [redacted] automorphism $\theta$ of $R^+$, $F(\theta d) = \theta F(d)$. A more useful characterization of this type of cluster method occurs in [13]. It states that a cluster method $F$ is monotone

equivariant if and only if

(1) The range of $Fd$ is always contained in the range of $d$,

and

(2) If $0 = h_0 < h_1 < \ldots < h_t$ is the range of $d$, then the sequence $TF(d)(h_0) \leq TF(d)(h_1) \leq \ldots \leq TF(d)(h_t)$ depends only upon the sequence $Td(h_0) < Td(h_1) < \ldots < Td(h_t)$ and is independent of the numerical values of the $h_i$'s.

Thus a monotone equivariant method may be regarded as a transformation of a sequence

$$R_0 \subset R_1 \subset \ldots \subset R_t = P \times P$$

of reflexive symmetric relations into a sequence

$$S_0 \subseteq S_1 \subseteq \ldots \subseteq S_t = P \times P$$

of relations which in most applications are taken to be equivalence relations. All of the algorithms we shall adopt are Type I algorithms in that for $i = 0,1,\ldots,t$, $S_i$ depends only upon $R_0,R_1,\ldots,R_i$.

Suppose now that one wants a measure of optimality that may be applied to Type I cluster methods. What properties should such a measure have? Though the following list is not intended to be complete, each of the conditions for such a measure does at least seem plausible.

(01) An optimality measure should be a mapping $\theta: C(P) \times C(P) \to T$, where $T = R^+$, $[0,1]$ or $[-1,1]$.

(02) $\theta(d,d') = \theta(d',d)$ for all $d,d' \in C(P)$.

(03) If $\theta(d,d') < \theta(d,d'')$, then in some readily interpretable sense, $d'$ should fit $d$ better than $d''$ does. Alternately, it would suffice if $-\theta$ had this property.

(04) $\theta$ should be monotone equivariant in the sense that if $\theta(d,d') < \theta(d,d'')$, and if $\theta$ is any order automorphism of $R^+$, then $\theta(\theta d,\theta d') < \theta(\theta d,\theta d'')$.

(05) Viewing a cluster method as a transformation of a sequence $R_0 \subset R_1 \subset \ldots \subset R_t = P \times P$ into a sequence $S_0 \subseteq S_1 \subseteq \ldots \subseteq S_t = P \times P$, it should not be possible to necessarily improve the value of 0 by setting $S_i = S_{i+1}$ for one or more values of i.

The exact meaning of axiom 05 will be clarified as it is applied in specific situations.

2. The optimality measures. A number of optimality measures will be considered. It will be convenient to briefly describe them in this section.

OP1. The cophenetic correlation coefficient. This is the ordinary product moment correlation between the input and output DC's. It was introduced by Sokal and Rohlf [22], and is probably the most commonly used measure of optimality. It should be noted, however, that some difficulties can occur when one applies the cophenetic correlation coefficient to input data having only ordinal significance. This is caused by the fact that this coefficient is not itself monotone equivariant, and does not behave well in the presence of a large number of ties. This was illustrated in [13]. p. 158 but it will be convenient to provide a similar illustration here. Let $P = \{a,b,c,e\}$, and consider the following set of binary attribute data:

| #attributes | $n_1$ | $n_2$ | $n_3$ | $n_4$ | $n_5$ | $n_6$ |
|---|---|---|---|---|---|---|
| a | 1 | 0 | 1 | 0 | 0 | 0 |
| b | 1 | 0 | 0 | 1 | 0 | 0 |
| c | 0 | 1 | 0 | 0 | 1 | 0 |
| e | 0 | 1 | 0 | 0 | 0 | 1 |

A dissimilarity measure will be defined from this table by using the simple matching coefficient. To obtain the value of $d(x,y)$ for $x,y \in P$, this amounts to counting up the number of attributes on which x,y disagree, and dividing by the total number of attributes. Three cases will be considered.

Case 1. $n_1 = n_2 = 1$; $n_3 = 3$, $n_4 = 1$, $n_5 = n_6 = 2$.

Case 2. $n_1 = n_2 = 3$, $n_3 = 3$, $n_4 = 1$, $n_5 = n_6 = 2$.

Case 3. $n_1 = n_2 = 8$; $n_3 = 31$, $n_4 = 1$, $n_5 = n_6 = 16$.

These lead respectively to the following three dissimilarity measures:

| $d_1$ | a | b | c | e |
|---|---|---|---|---|
| a | 0 | .4 | .7 | .7 |
| b | | 0 | .5 | .5 |
| c | | | 0 | .4 |

| $d_2$ | a | b | c | e |
|---|---|---|---|---|
| a | 0 | .29 | .79 | .79 |
| b | | 0 | .64 | .64 |
| c | | | 0 | .29 |

| $d_3$ | a | b | c | e |
|---|---|---|---|---|
| a | 0 | .40 | .79 | .79 |
| b | | 0 | .41 | .41 |
| c | | | 0 | .40 |

These 3 dissimilarity measures are equivalent in terms of their rank orderings, so a monotone equivariant cluster method will necessarily produce an equivalent output for each of them. Let us consider the following candidates for output DC's:

| $d_4$ | a | b | c | e |
|---|---|---|---|---|
| a | 0 | k | k | k |
| b | | 0 | j | j |
| c | | | 0 | h |

| $d_5$ | a | b | c | e |
|---|---|---|---|---|
| a | 0 | h | j | j |
| b | | 0 | j | j |
| c | | | 0 | h |

Here in Case 1, h = .4, j = .5, k = .7; in Case 2, h = .29, j = .64, k = .79, and in Case 3, h = .40, j = .41 and k = .79. Let us now examine the values of the cophenetic correlation coefficient for these 2 outputs.

| | $d_4$ | $d_5$ |
|---|---|---|
| Case 1 | .59 | .78 |
| Case 2 | .55 | .96 |
| Case 3 | .70 | .52 |

Notice that in Case 3, the cophenetic correlation coefficient predicts that $d_4$ is superior to $d_5$; yet no Type I cluster method would ever produce $d_4$ since the first level of input data leads naturally to the classification {a,b}, {c,e} while $d_4$ reflects the classification {a}, {b}, {c,e}. Apart from the above problems, the values of an input DC are not independent, so it is difficult to

provide any statistical significance to this coefficient. Indeed, as was pointed out by Jardine and Sibson [16], p. 107, "...justification for its use rests on its practical utility rather than on any resemblance to statistical correlation measures."

Next we consider some optimality measures that are based upon distance measures. In what follows, the vector $(x_1,x_2,\ldots,x_k)$ will denote the values of the input DC, and the vector $(y_1,y_2,\ldots,y_k)$ the values of the output of some cluster method under consideration. We then have

OP2. $[\Sigma(x_i - y_i)^2]^{.5}$

OP3. $[\Sigma(x_i - y_i)^2]^{.5}/(x_i^2)^{.5}$

OP4. $[\Sigma(1 - (y_i/x_i))^2]^{.5}$

OP5. $\Sigma|x_i - y_i|$ .

Measures OP2 and OP5 are both suggested by Jardine and Sibson [16], p.103. The remaining measures are in Kruskal [17], pp. 307-8 for use as stress measures in multidimensional scaling. They are each measures of the distance between the input and output DC's, and from our point of view each has as a major defect the fact that it is not monotone equivariant, and consequently should not be used as an optimality measure with ordinal data.

Hubert and Baker [9] introduce a measure that is based upon the Goodman-Kruskal correlation coefficient [3]. To see how it goes, we consider all ordered pairs of unordered pairs of elements of P. Assuming that d is a dissimilarity measure on P and that F is a cluster method, the pair $(\{a,b\},\{c,e\})$ is called <u>consistent</u> if $d(a,b) < d(c,e)$ and $Fd(a,b) < Fd(c,e)$; it is called <u>inconsistent</u> if $d(a,b) < d(c,e)$ and $Fd(a,b) > Fd(c,e)$. One then lets $S^+$ = the number of consistent pairs, $S^-$ the number of inconsistent pairs, and takes

$\gamma = \frac{S^+ - S^-}{S^+ - S^-}$ . This assumes that there are no ties in the values of the input data provided by d. In the presence of T ties in d, one can argue that $\gamma$ must lie between $\frac{S^+ - S^- - T}{S^+ + S^- + T}$ and $\frac{S^+ - S^- + T}{S^+ + S^- + T}$ . This amounts to considering the two extreme possibilities: that a tie always leads to an inconsistent pair, and that a tie always leads to a consistent pair. With this thoughtin mind, we shall use the notation

$$\text{OP7.} \quad \gamma_L = \frac{S^+ - S^- - T}{S^+ + S^- + T}$$

$$\text{OP8.} \quad \gamma_U = \frac{S^+ - S^- + T}{S^+ + S^- + T}$$

and take $\gamma = [\gamma_L, \gamma_U]$. For further details of the derivation of $\gamma$ as well as for a discussion of some of its statistical properties, the reader is referred to [9]. At first glance the $\gamma$ -coefficient seems extremely attractive. First of all, it is monotone equivariant. Secondly, it has an interesting probabilistic interpretation. Finally, it seems intuitively clear that higher values of $\gamma$ should lead to the conclusion that the corresponding cluster methods are better reflections of the input data. But more careful reflection shows that this measure has a danger associated with it. The danger is caused by the fact that it ignores pairs $(\{a,b\},\{c,e\})$ where $d(a,b) < d(c,e)$ and $Fd(a,b) = Fd(c,e)$. This can best be illustrated with a concrete example that we take from [9], p.90:

Here $P = \{1,2,3,4,5,6\}$, and d is given by

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | | 3 | 4 | 14 | 1 | 8 |
| 2 | | | 5 | 9 | 11 | 12 |
| 3 | | | | 13 | 2 | 10 |
| 4 | | | | | 6 | 7 |
| 5 | | | | | | 15 |

The complete linkage output for this is:

| Level | Nontrivial clusters |
|---|---|
| 1 | {1,5} |
| 4 | {1,3,5} |
| 7 | {1,3,5}, {4,6} |
| 11 | {1,2,3,5}, {4,6} |
| 15 | {1,2,3,4,5,6} |

The value of $\gamma$ for the above output is 0.78. By discarding levels of output, this value of $\gamma$ can be "improved" as follows:

| Output levels | $\gamma$ |
|---|---|
| 1,4,11,15 | 0.83 |
| 1,4,7,15 | 0.84 |
| 1,4,15 | 0.95 |
| 1,15 | 1.00 |

Indeed, $\gamma$ will equal 1.00 for any input DC d, if the most similar pair of objects is clustered at level 1 and all other pairs at level 2. It would seem that $\gamma$ tends to favor cluster methods that produce few output clusters. Apart from this, $\gamma$ is clearly not going to give much information in the presence of a large number of ties. This sort of problem has led us to modify the definition of $\gamma$. Specifically, one takes $(\{a,b\},\{c,e\})$ to be a consistent pair if either $d(a,b) < d(c,e)$ and $Fd(a,b) \leq Fd(c,e)$ or $d(a,b) = d(c,e)$ and $Fd(a,b) = Fd(c,e)$[1]; one takes it to be inconsistent if $d(a,b) < d(c,e)$ and $Fd(a,b) > Fd(c,e)$ or if $d(a,b) = d(c,e)$ and $Fd(a,b) > Fd(c,e)$. This modification reflects the fact that if $(c,e)$ is clustered at level $h$, and if $(a,b)$ is more similar or equally similar, then it too should be clustered at level $h$. One defines $S^+$ and $S^-$ as was done earlier with

OP6. $\gamma' = \dfrac{S^+ - S^-}{S^+ + S^-}$ .

Note, however that $\gamma'$ still suffers from the defect that it can be "improved" by removing desirable cluster levels. (See the table at the end of the discussion of OP11).

[1] Each pair $(\{a,b\},\{c,e\})$ to be counted only once.

This brings us to

OP9. Kendall's Tau-correlation coefficient.

OP10. Spearman's Rho-correlation coefficient.

These are both rank order correlation coefficients, and are consequently monotone equivariant. Like the cophenetic correlation coefficient, they tend to behave rather poorly in the presence of a large number of ties. Furthermore, since the underlying statistical assumptions that validate their use as correlation coefficients are simply not present, no statistical significance should be attached to their use. They should be viewed as ad hoc measures of optimality; nonetheless, they are certainly as useful as the cophenetic correlation coefficient - especially for ordinal input data. Their most serious flaw is that they rank order both the input and the output dissimilarity measures. For that reason, the following outputs would be treated as being identical:

| Level | | | | |
|---|---|---|---|---|
| $h_1$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ |
| E | P × P | P × P | P × P | P × P |
| E | E | P × P | P × P | P × P |
| E | E | E | P × P | P × P |
| E | E | E | E | P × P |

Here E is some fixed equivalence relation on the set P.

We now introduce an optimality measure that directly determines how well a Type I method performs. One should first note that a Type I cluster method tries to group $\{a,b\}$ into a single cluster at level h if $d(a,b) \leq h$, and tries to leave it ungrouped if $d(a,b) > h$. Suppose the input DC d on P has range $h_1,h_2,\ldots,h_t$ where $h_1 < h_2 < \ldots < h_t$. Suppose further that $d(a,b) = h_i$ and $Fd(a,b) = h_{i+k}$ $(k > 0)$. One then has

| Level $h_j$ | Classification of (a,b) | Explanation |
|---|---|---|
| $j < i$ | Proper | $d(a,b) > h_j$<br>$Fd(a,b) > h_j$ |
| $i + k > j \geq i$ | Improper | $d(a,b) \leq h_j$<br>$Fd(a,b) > h_j$ |
| $j \geq k + i$ | Proper | $d(a,b) < h_j$<br>$Fd(a,b) \leq h_j$ |

Thus there are $k$ levels at which $(a,b)$ is not properly classified. If $d(a,b) = h_i$ and $Fd(a,b) = h_{i-k}$, a similar argument shows that $(a,b)$ is not properly classified at $k$ levels. But this makes it extremely easy to count up the number of points at each level that are not properly classified, this sum being taken over all possible levels. If the input DC d has as its range $h_1, h_2, \ldots, h_t$ $(h_1 < h_2 < \ldots < h_t)$ one just replaces $h_i$ with $i$ in both the input and output *DC's (recalling that the range of the output is* contained in the range of the input DC and then computes

$$\Sigma |d(a,b) - Fd(a,b)| \quad (a,b \in P).$$

This may be normalized by dividing by the total number of points under consideration. If $|P| = p$, this is simply $tp(p-1)/2$, where $t$ is the number of levels in the range of the input data. Incidentally, this leads to a useful interpretation of the measure. If both the selection of points and levels are equally probable, then this measure provides the probability that a random point be improperly classified at a randomly chosen level. In addition to this, it has a number of pleasant properties. It is monotone equivariant. It directly measures the performance of a Type I cluster method. Finally, it cannot be "improved" by discarding desirable clusters. This coefficient will be referred to as OP11. Incidentally, OP6, OP7, OP8 are unique in that they can be improved by discarding

cluster levels of the output. To illustrate this, let us again consider the example that was presented in connection with the Goodman-Kruskal coefficient $\gamma$:

| Output levels | OP1 | OP2 | OP3 | OP4 | OP5 | OP6 | OP9 | OP10 | OP11 |
|---|---|---|---|---|---|---|---|---|---|
| 1,4,7, 11,15 | 0.77 | 17.6 | 0.50 | 3.65 | 49 | 0.85 | 0.65 | 0.77 | 0.22 |
| 1,4,11, 15 | 0.77 | 18.0 | 0.51 | 3.70 | 53 | 0.89 | 0.68 | 0.78 | 0.24 |
| 1,4,7, 15 | 0.67 | 21.7 | 0.62 | 4.99 | 61 | 0.92 | 0.57 | 0.66 | 0.27 |
| 1,4,15 | 0.66 | 23.1 | 0.66 | 5.12 | 69 | 0.98 | 0.57 | 0.66 | 0.31 |
| 1,15 | 0.43 | 28.6 | 0.81 | 8.66 | 91 | 1.00 | 0.37 | 0.43 | 0.40 |

There is a final optimality measure that will be introduced. It is of a different character than those that have been considered earlier in that it tells how well a given equivalence relation fits an input DC, as opposed to determining how well the output DC of a cluster method fits the input DC. It was introduced by Hubert ([6],[7]). To see how it goes, consider a set P equipped with a DC d. If $d(a,b) < d(c,e)$, $Fd(c,e) \leq h$ and $Fd(a,b) > h$, one says that there is a discrepency at level h. Let $E = TF(d)(h)$. Suppose E has clusters $C_1, C_2, \ldots, C_k$ and that $|C_i| = r_i$. One then defines $\alpha_h$ to be the quotient of the number of discrepencies over the total possible number of discrepencies. If $|P| = p$, this denominator is given by

$$[\Sigma \frac{r_i(r_i - 1)}{2}][\frac{p(p-1)}{2} - \Sigma \frac{r_i(r_i - 1)}{2}] .$$

Properties enjoyed by $\alpha_h$ include the following:

1. It is monotone equivariant.

2. It deals with individual equivalence relations.

3. A low value of $\alpha_h$ may readily be interpreted as meaning that the output at level h provides a good fit to the input data.

4. $\alpha_h$ ignores ties in the input data, and for that reason can give misleading results in the presence of a large number of ties. To see this, take $P = \{0_1,0_2,\ldots,0_{100}\}$ with d defined by $d(0_1,0_2) = 1$, $d(0_i,0_j) = 2$ for all other values of i and j $(i \neq j)$. Then if E has $\{0_1,0_2,0_3\}$ as its only nontrivial class, then $\alpha_h(E) = 0$ despite the fact that E does not provide a particularly close fit to d.

This is the last of the optimality measures that we shall consider. For the reader's convenience, these measures are summarized in Table 1 of the Appendix.

3. A classification of optimality measures. At this point a total of 11 global optimality measures have been introduced - each of them tells us something about the goodness of fit of the output DC of a cluster method to the input DC. In order to make some sense out of these measures, and to make an intelligent choice among them, let us compare them on some actual data. The input data is that described in [15], pp. 11-16. The idea is to start with the binary data contained in [15], Tables 1-5, introduce a 5% random error in each character, add 6 random characters, and finally discard 10% of the resulting characters. This data is then converted to a dissimilarity coefficient in each of the following ways:

| | | |
|---|---|---|
| DC1 | (b + c)/n | Simple matching coefficient. |
| DC4 | (b + c + d)/n | Coefficient of Russell & Rao. |
| DC10 | bc/(ad + bc) | Yule's coefficient. |

Note: For a pair J,K of objects,

d = the number of shared 0's;
a = the number of shared 1's'
b = number of characters for which J has 1, K has 0;
c = number of characters for which K has 1, J has 0;
n = a + b + c + d.

For each data set, 10 trials were performed, $u = .5$ clustering was applied and all 11 optimality measures used. In addition to this, 10 trials were performed using 25 random characters on a 10 element set. Where needed each optimality measure was then converted to a dissimilarity measure. The similarity between the predictions made by the optimality measures was then computed in two different ways:

1. Using the product-moment correlation;
2. Using the Goodman-Kruskal rank-order correlation.

   (Note. This amounts to using OP7 with $T = 0$, since there are no ties in the input data).

These correlations were then converted to dissimilarity measures and $u = .5$-clustering applied. The cluster outputs appear in the Appendix as Tables 2-7, and in summary form in Table 8. A brief examination of Table 8 suggests that the following grouping of optimality measures might be significant:

Group I. OP2, OP3, OP4, OP5

Group II. OP6, OP7

Group III. OP1, OP9, OP10, OP11.

In a sense this is hardly surprising. Group I contains the distance measures,, OP6 is a minor modification of OP7, and Group III contains the 3 correlation measures as well as OP11. This leaves OP8 unclassified. We choose to place it in Group II because its very definition shows it to be closely related to OP7.

In what follows, we shall take as representative optimality measures the following:

OP2 from Group I,
OP7-8 from Group II
OP1 from Group III
OP11 and OP12 because of their desirable properties.

A classification of optimality measures was also undertaken by Gower and Banfield [4] with somewhat similar conclusions. These authors used a product moment correlation to compare the various measures, and followed this by a principal coordinates analysis. Though their choice of optimality measures differed somewhat from ours, they did use OP1, OP2, OP3 and OP9. Despite this, the reader is urged to use some caution in interpreting the results that we have just announced, as they could be somewhat dependent upon the size of the data sets (see [4] pp. 355-357).

4. The cluster techniques. Having arrived at some reasonable optimality measures, it is now appropriate to apply them to some actual data. The cluster techniques we shall use have all been described in [12], but for the readers convenience, we shall briefly mention them here again. The input data should be regarded as a DC on a finite set P of objects to be classified. Let $h_0 < h_1 < \ldots < h_t$ be the range of d. The techniques are all agglomerative. At level $h_0$, J and K are grouped together if $d(J,K) = h_0$; the transitive closure of the resulting relation is then formed. At level $h_{i+1}$, the input data consists of d and the clusters that were formed at level $h_i$. The techniques differ only in the criterion they employ for merging 2 clusters at level $h_{i+1}$. A link between clusters $J,K$ at that level consists of a pair of objects $J,K$ with $J \in J$, $K \in K$, and $d(J,K) \leq h_{i+1}$. There are two types of algorithms:

u-clustering $(0 \leq u \leq 1)$. This merges $J,K$ if at least a fraction u of all possible links have been made between them.

(u,v)-clustering $(0 < u,v \leq 1)$. Here $J,K$ are merged if at least a fraction u of the members of $J$ are each linked to at least a fraction v of the elements of $K$, and similarly, at least u of the elements of $K$ are each linked to at least v of the elements of $J$.

Finally, in the spirit of comparison, we shall employ the UPGMA algorithm as described in [21], pp. 230-234. This cluster technique has enjoyed considerable popularity with numerical taxonomists, and tends to produce useful cluster outputs for their purposes. Despite this, it is not monotone equivariant, and also suffers from the defect that its output is dependent on the labeling of the input data. Because such techniques should not be used on input data having only ordinal significance, it will be interesting to compare the effectiveness of UPGMA with that of some of the u- and uv-techniques.

5. Some numerical examples. Several numerical examples will be considered. On each example, we shall apply single linkage, complete linkage, u-clustering (u = .3, .5, .7), uv-clustering (uv = (.2,.4), (.2,.6), (.4,.6)) as well as the UPGMA algorithm. The idea will be to comapre the optimality of these cluster techniques, as measured by OP1, OP2, OP7-8 and OP11. The optimality measures will be viewed as simple numerical measures of how well a cluster method has performed; no attempt will be made to attach any statistical significance to their predictions.

Example 1. The data consists of Holzinger and Harman's eight psychological tests [5] and appears in Table 9, with the various cluster outputs appearing in Table 10. Notice that apart from the levels at which the clusters appear, there are only 3 distinct outputs. The various optimality measures occur in Table 15. Notice that these measures do indeed produce different verdicts. OP1 rates UPGMA as best with uv = (.2,.4) a close second. By OP2, UPGMA is best with u = .5 second best. On the other hand, the Goodman-Kruskal coefficient tells us that single linkage, u = .3, u = .5, uv = (.2,.4) and UPGMA are all equally good. Finally, OP11, which in a certain sense

directly measures how many objects have been misclassified, puts uv = (.2,.6) best and uv = (.4,.6) second best. The only agreement between these measures is that complete linkage is the worst choice of all of those considered.

Example 2. Here we consider some data from Cooley and Lohnes [2], pp. 133-134. An explanation of the construction of the data table can be found in [7]. The actual data is reproduced here as Table 11 with the cluster outputs in Table 12. Here again the optimality measures (See Table 15) provide a mixed verdict. By OP1, UPGMA and uv = (.4,.6) are tied for first place, while by OP2, UPGMA is best with uv = (.2,.6), uv = (.2,.4) and uv = (.4,.6) close behind. The Goodman-Kruskal coefficient rates UPGMA, the 3 uv methods and u = .3, u = .7 as being equally good. Finally, OP11 puts uv = (.2,.6) and uv = (.4,.6) first and second best. Again, complete linkage is rated the worst choice by all of the optimality measures. Incidentally, the outputs rated best by Goodman-Kruskal coincide with those announced by Hubert [7], p.57 for his "objective function hierarchy".

Example 3. This represents 13 diagnostic types that were discussed by Overall [19]. His categorization produces the partition

123,46789-10,11-12,5,13

which appears in all cluster outputs with the exception complete linkage (See Table 13). Here uv = (.2,.4) is rated best by OP1, UPGMA by OP2, uv = (.4,.6) by Goodman-Kruskal and u = .5 by OP11. The optimality measures all agree that single linkage and complete linkage clustering are the worst choices of those considered. The reader might compare these results with a study made by Hubert [8].

Example 4. This involves quantifying the confusability among 16 consonant phonomes as reported by Miller and Nicely [18]. The acutal data we used appears as matrix (b) in Hubert and Baker [9], p. 105. Here it is convenient to follow the notation of [9], and let the first 16 integers denote the consonant phonomes ordered as follows: p, t, k, f, θ, s, ʃ, b, g, v, ð, ʒ, d, z, m, n. The various cluster outputs are almost identical (Table 14), but again the optimality measures produce a somewhat mixed verdict. OP1 rates as best u = .3, u = .7, uv = (.4,.6) and UPGMA; OP2 puts UPGMA best with uv = (.2,.6), second best; OP7-8 rate u = .3, .u = .7, uv = (.2,.6), uv = (.4,.6) and UPGMA as best; finally, OP11 says that the best outputs are those produced by uv = (.2,.6) and uv = (.4,.6). As in Example 3, the optimality measures all conclude that single linkage and complete linkage are the worst choices for cluster techniques of those considered. Notice that the outputs for u = .7, uv = (.2,.6), uv = (.4,.6) and UPGMA clustering are all equivalent. The clusters that occur at the first level at which no singleton clusters remain (level 9) all occur in a fairly natural manner and may be given reasonable interpretations. (See [20], Table 2, p. 105). They are:

| | |
|---|---|
| 123 | unvoiced stops |
| 45 | front unvoiced fricatives |
| 67 | back unvoiced fricatives |
| 8-11-12 | front voiced consonants |
| 9-10 | back voiced stops |
| 13-14 | back voiced fricatives |
| 15-16 | nasals |

The level 13 partition consists of the unvoiced consonants, the voiced consonants, and the nasals.

6. Summary. Some of the more commonly used optimality measures for cluster techniques were introduced and their properties discussed with a view

toward their possible utility for monotone equivariant cluster techniques. They were grouped naturally into 3 classes and a representative chosen from each class. A new optimality measure was also introduced; it directly measures the effectiveness of a certain type of monotone equivariant cluster methods. These measures were then applied to several numerical examples to compare the outputs of single linkage, complete linkage and UPGMA with certain monotone equivariant cluster techniques that have recently been introduced by the author. In these examples, neither single linkage nor complete linkage performed well; the best overall performance seemed to be by UPGMA, $uv = (.2,.6)$ and $uv = (.4,.6)$. When faced with ordinal input data, this might be taken as evidence that some of the uv-techniques could prove useful.

7. Acknowledgement. The computer programs needed for this study were written in APL by Zachary Smith, and the data was processed on the University of Massachusetts Control Data Corporation CYBER 70 computer.

Table 1.

| Symbol | Name | Formula (See Note 4) |
|---|---|---|
| OP1 | Cophenetic Correlation Coefficient | Product moment correlation |
| OP2 | Euclidean distance | $[\Sigma(x_i - y_i)^2]^{.5}$ |
| OP3 | | $[\Sigma(x_i - y_i)^2]^{.5}/(\Sigma x_i^2)^{.5}$ |
| OP4 | | $[\Sigma(1 - (y_i/x_i))^2]^{.5}$ |
| OP5 | Manhattan distance | $\Sigma\|x_i - y_i\|$ |
| OP6 | Modified Goodman-Kruskal | $(S^+-S^-)/(S^++S^-)$ (See Note 1) |
| OP7 | Goodman-Kruskal | $(S^+-S^--T)/(S^++S^-+T)$ (See Note 2) |
| OP8 | Goodman-Kruskal | $(S^+-S^-+T)/(S^++S^-+T)$ (See Note 2) |
| OP9 | Kendall's Tau | |
| OP10 | Spearman's Rho | |
| OP11 | | $(\Sigma\|x_i - y_i\|)/[kp(p-1)/2]$ where $\|P\| = p$ and $k$ = no. splitting levels. (See Note 3) |

Note 1. For definitions of $S^+$ and $S^-$, see p.8

Note 2. For definitions of $S^+$, $S^-$, T, see p.6-7

Note 3. The input data is first rank ordered. See p. 10

Note 4. In the table $(x_1, x_2, \ldots, x_t)$ denotes the values of the input DC d, and $(y_1, y_2, \ldots, y_t)$ the corresponding values of the output DC F(d).

Table 2. Cluster output for optimality measures with data from Table 1 of [15].

| Level | Product-moment | Goodman-Kruskal |
|---|---|---|
| 1 | 6 7 | 6 7 |
| 2 | 2 5, 6 7 | 2 5, 6 7 |
| 3 | 2 5, 6 7, 9 10 | 2 5, 3 4, 6 7 |
| 4 | 2 3 5, 6 7, 9 10 | 2 5, 3 4, 6 7, 9 10 |
| 5 | 2 3 4 5, 6 7, 9 10 | 2 3 4 5, 6 7, 9 10 |
| 6 | 2 3 4 5, 6 7, 1 9 10 | 1 11, 2 3 4 5, 6 7, 9 10 |
| 7 | 2 3 4 5, 1 6 7 9 10 | 1 11, 2 3 4 5, 6 7 9 10 |
| 8 | 2 3 4 5, 1 6 7 9 10 11 | 1 8 11, 2 3 4 5, 6 7 9 10 |
| 9 | 2 3 4 5, 1 6 7 8 9 10 11 | 2 3 4 5, 1 6 7 8 9 10 11 |

Table 3. Cluster output for optimality measures with data from Table 6 of [15].

| Level | Product-moment | Goodman-Kruskal |
|---|---|---|
| 1 | 2 5 | 2 3 |
| 2 | 2 5, 3 4 | 2 3, 6 7 |
| 3 | 2 5, 3 4, 6 7 | 2 3 4, 6 7 |
| 4 | 2 3 4 5, 6 7 | 2 3 4 5, 6 7 |
| 5 | 2 3 4 5, 6 7, 9 10 | 2 3 4 5, 6 7, 1 9 |
| 6 | 2 3 4 5, 6 7, 1 9 10 | 2 3 4 5, 6 7, 1 9 10 |
| 7 | 2 3 4 5, 6 7, 1 8 9 10 | 2 3 4 5 8, 6 7, 1 9 10 |
| 8 | 2 3 4 5, 1 6 7 8 9 10 | 2 3 4 5 8, 6 7, 1 9 10 11 |
| 9 | 2 3 4 5 11, 1 6 7 8 9 10 | 2 3 4 5 8, 1 6 7 9 10 11 |

Table 4. Cluster output for optimality measures based on data from Table 7 of [15].

| Level | Product-moment | Goodman-Kruskal |
|---|---|---|
| 1 | 2 5 | 3 4 |
| 2 | 2 5, 9 10 | 2 5, 3 4 |
| 3 | 2 5, 6 7, 9 10 | 2 5, 3 4, 9 10 |
| 4 | 2 3 5, 6 7, 9 10 | 2 5, 3 4, 6 7, 9 10 |
| 5 | 2 3 4 5, 6 7, 9 10 | 2 3 4 5, 6 7, 9 10 |
| 6 | 1 9 10, 2 3 4 5, 6 7 | 1 11, 2 3 4 5, 6 7, 9 10 |
| 7 | 2 3 4 5, 1 9 10, 6 7 8 | 2 3 4 5, 6 7, 1 9 10 11 |
| 8 | 2 3 4 5, 1 6 7 8 9 10 | 2 3 4 5 8, 6 7, 1 9 10 11 |
| 9 | 2 3 4 5, 1 6 7 8 9 10 11 | 2 3 4 5 8, 1 6 7 9 10 11 |

Table 5. Cluster output for optimality measures based on data from Table 8 of [15].

| Level | Product-moment | Goodman-Kruskal |
|---|---|---|
| 1 | 2 5 | 2 5 |
| 2 | 2 5, 6 7 | 2 5, 6 7 |
| 3 | 2 5, 3 4, 6 7 | 2 5, 3 4, 6 7 |
| 4 | 2 5, 3 4, 6 7, 9 10 | 2 5, 3 4, 6 7, 9 10 |
| 5 | 2 3 4 5, 6 7, 9 10 | 2 3 4 5, 6 7, 9 10 |
| 6 | 2 3 4 5, 6 7, 1 9 10 | 2 3 4 5, 6 7, 1 9 10 |
| 7 | 2 3 4 5, 6 7, 1 9 10 11 | 2 3 4 5, 6 7, 1 9 10 11 |
| 8 | 2 3 4 5, 6 7 8, 1 9 10 11 | 2 3 4 5, 6 7 8, 1 9 10 11 |
| 9 | 2 3 4 5, 1 6 7 8 9 10 11 | 2 3 4 5, 1 6 7 8 9 10 11 |

Table 6. Cluster output for optimality measures based on data from Table 9 of [15].

| Level | Product-moment | Goodman-Kruskal |
|---|---|---|
| 1 | 6 7 | 6 7 |
| 2 | 2 5, 6 7 | 3 4, 6 7 |
| 3 | 2 5, 6 7, 9 10 | 2 5, 3 4, 6 7 |
| 4 | 2 3 5, 6 7, 9 10 | 2 5, 3 4, 6 7, 9 10 |
| 5 | 2 3 4 5, 6 7, 9 10 | 2 5, 3 4, 6 7, 1 9 10 |
| 6 | 2 3 4 5, 6 7, 1 9 10 | 2 3 4 5, 6 7, 1 9 10 |
| 7 | 2 3 4 5, 6 7, 1 9 10 11 | 2 3 4 5, 6 7, 1 9 10 11 |
| 8 | 2 3 4 5 8, 6 7,1 9 10 11 | 2 3 4 5, 6 7, 1 8 9 10 11 |
| 9 | 2 3 4 5 8, 1 6 7 9 10 11 | 2 3 4 5, 1 6 7 8 9 10 11 |

Table 7. Cluster output for optimality measures based on random data.

| Level | Product-moment | Goodman-Kruskal |
|---|---|---|
| 1 | 2 5 | 9 10 |
| 2 | 2 5, 9 10 | 2 5, 9 10 |
| 3 | 2 3 5, 9 10 | 2 3 5, 9 10 |
| 4 | 2 3 5, 1 9 10 | 2 3 5, 6 7, 9 10 |
| 5 | 2 3 5, 6 7, 1 9 10 | 2 3 5, 6 7, 1 9 10 |
| 6 | 2 3 4 5, 6 7, 1 9 10 | 2 3 4 5,6 7, 1 9 10 |
| 7 | 2 3 4 5, 6 7, 1 8 9 10 | 2 3 4 5, 6 7, 1 8 9 10 |
| 8 | 2 3 4 5, 1 6 7 8 9 10 | 2 3 4 5, 1 6 7 8 9 10 |
| 9 | 2 3 4 5 11, 1 6 7 8 9 10 | 2 3 4 5 11, 1 6 7 8 9 10 |

Table 8. Number of occurrences of clusters in Tables 2 through 7.

| No. Occurrences | Cluster | No. Occurrences | Cluster |
|---|---|---|---|
| 12 | 6 7 | 3 | 6 7 8 |
| | 2 3 4 5 | | 2 3 4 5 8 |
| 11 | 2 5 | | 2 3 4 5 11 |
| | 9 10 | | 1 8 9 10 |
| 10 | 1 9 10 | 2 | 1 11 |
| 6 | 3 4 | 1 | 2 3 |
| | 1 9 10 11 | | 1 9 |
| | 1 6 7 8 9 10 11 | | 2 3 4 |
| 5 | 2 3 5 | | 6 7 9 10 |
| 4 | 1 6 7 8 9 10 | | 1 8 11 |
| | 1 6 7 9 10 11 | | 1 6 7 9 10 |
| | | | 1 8 9 10 11 |

Table 9. Holzinger and Harman's proximity values for eight psychological tests.

| Object pair | DC | Object pair | DC | Object pair | DC |
|---|---|---|---|---|---|
| {6,7} | .278 | {4,7} | .665 | {2,5} | .715 |
| {5,7} | .344 | {1,6} | .665 | {3,6} | .732 |
| {5,6} | .378 | {1,8} | .668 | {3,5} | .753 |
| {7,8} | .381 | {4,6} | .673 | {2,6} | .766 |
| {5,8} | .422 | {1,5} | .679 | {2,4} | .770 |
| {6,8} | .473 | {1,2} | .682 | {4,5} | .773 |
| {1,4} | .532 | {2,3} | .683 | {3,7} | .777 |
| {1,3} | .597 | {3,4} | .695 | {2,7} | .843 |
| {4,8} | .609 | {1,7} | .696 | {2,8} | .843 |
| {3,8} | .618 | | | | |

Table 10. Cluster outputs for data in Table 9.

| Clusters | α | Output DC value for given method UPGMA | single | u=.3 | u=.5 | uv=(.2,.4) |
|---|---|---|---|---|---|---|
| 67 | 0 | .278 | .278 | .278 | .278 | .278 |
| 567 | 0 | .361 | .344 | .344 | .344 | .344 |
| 5678 | 0 | .425 | .381 | .381 | .422 | .422 |
| 14,5678 | 0 | .532 | .532 | .532 | .532 | .532 |
| 134,5678 | .053 | .646 | .597 | .597 | .597 | .597 |
| 1345678 | .136 | .692 | .609 | .665 | .673 | .665 |
| 12345678 | | .757 | .682 | .715 | .766 | .715 |
| | | Output DC for given method u=.7 | uv=(.2,.6) | | uv=(.4,.6) | |
| 67 | 0 | .278 | .278 | | .278 | |
| 567 | 0 | .378 | .378 | | .278 | |
| 5678 | 0 | .473 | .422 | | .422 | |
| 14,5678 | 0 | .532 | .532 | | .532 | |
| 145678 | .133 | .679 | .673 | | .673 | |
| 145678,23 | .141 | .683 | .683 | | .683 | |
| 12345678 | | .770 | .753 | | .766 | |
| | | Output DC for complete linkage | | | | |
| 67 | 0 | .278 | | | | |
| 567 | 0 | .378 | | | | |
| 5678 | 0 | .473 | | | | |
| 14,5678 | 0 | .532 | | | | |
| 14,23,5678 | .056 | .683 | | | | |
| 1234,5678 | .172 | .770 | | | | |
| 12345678 | | .843 | | | | |

Table 11. Dissimilarity values for data from Cooley and Lohnes

| Object pair | DC | Object pair | DC | Object pair | DC |
|---|---|---|---|---|---|
| {1,3} | .013 | {2,7} | .123 | {1,7} | .403 |
| {5,8} | .016 | {5,7} | .152 | {2,5} | .509 |
| {1,8} | .020 | {4,7} | .165 | {4,5} | .511 |
| {3,8} | .028 | {3,6} | .193 | {4,8} | .653 |
| {5,6} | .038 | {1,6} | .196 | {2,8} | .669 |
| {6,7} | .050 | {7,8} | .257 | {3,4} | .815 |
| {1,5} | .069 | {2,6} | .307 | {1,4} | .832 |
| {2,4} | .070 | {4,6} | .329 | {2,3} | .840 |
| {3,5} | .073 | {3,7} | .392 | {1,2} | .860 |
| {6,8} | .097 | | | | |

Table 12. Cluster outputs for data from Table 11.

| | | Output DC value for given method | | | | | |
|---|---|---|---|---|---|---|---|
| Clusters | $\alpha$ | UPGMA | u=.3 | u=.7 | .2,.4 | .2,.6 | .4,.6 |
| 13 | 0 | .013 | .013 | .013 | .013 | .013 | .013 |
| 13,58 | 0 | .016 | .016 | .016 | .016 | .016 | .016 |
| 1358 | .038 | .048 | .028 | .050 | .020 | .050 | .050 |
| 1358,67 | .027 | .050 | .050 | .069 | .050 | .069 | .069 |
| 1358,24,67 | .025 | .070 | .070 | .070 | .070 | .070 | .070 |
| 135678,24 | .078 | .216 | .152 | .257 | .097 | .193 | .257 |
| 12345678 | | .551 | .329 | .815 | .509 | .653 | .653 |
| | | Output DC value for single linkage | | | | | |
| 13 | 0 | .013 | | | | | |
| 13,58 | 0 | .016 | | | | | |
| 1358 | .038 | .020 | | | | | |
| 13568 | .083 | .038 | | | | | |
| 135678 | .118 | .050 | | | | | |
| 135678,24 | .078 | .070 | | | | | |
| 12345678 | | .123 | | | | | |

Table 12 (continued)

| Clusters | α | Output DC value for u=. 5 |
|---|---|---|
| 13 | 0 | .013 |
| 13,58 | 0 | .016 |
| 1358 | .038 | .028 |
| 1358,67 | .027 | .050 |
| 1358,24,67 | .025 | .070 |
| 1358,2467 | .109 | .165 |
| 12345678 | | .403 |
| | | Output DC for complete linkage |
| 13 | 0 | .013 |
| 13,58 | 0 | .016 |
| 13,58,67 | .040 | .050 |
| 13,24,58,67 | .073 | .070 |
| 1358,24,67 | .025 | .073 |
| 1358,2467 | .109 | .329 |
| 12345678 | | .860 |

Table 13. Portion of Cluster Outputs for Overall's 13 diagnostic types.

| | α | Output DC value for given method | | | | |
|---|---|---|---|---|---|---|
| | | single | u=.7 | uv=.2,.4 | .2,.6 | .4,.6 |
| 123,4789-10 11-12 | .020 | 3.58 | 10.84 | 7.6 | 7.6 | 7.86 |
| 123,46789-10, 11-12 | .022 | 5.18 | 13.14 | 7.81 | 9.24 | 9.24 |
| 12346789-10 11-12 | .097 | 10.57 | 22.91 | 13.55 | 13.69 | 14.55 |
| | | Output DC value for given method | | | | |
| | | u=.5 | UPGMA | | | |
| 123,4789-10, 11-12 | .020 | 7.60 | 8.26 | | | |
| 123,46789-10, 11-12 | .022 | 9.24 | 10.70 | | | |
| 123,456789-10, 11-12 | .076 | 17.64 | 19.51 | | | |
| | | Output DC value for u=.3 | | | | |
| 123,4-10, 6789,11-12 | .026 | 5.18 | | | | |
| 123,46789-10, 11-12 | .022 | 7.6 | | | | |
| 12346789-10, 11-12 | .097 | 13.55 | | | | |
| | | Output DC value for complete linkage | | | | |
| 123,4789-10, 11-12 | .020 | 12.97 | | | | |
| 123,4789-10, 56,11-12 | .030 | 13.84 | | | | |
| 123,456789-10, 11-12 | .075 | 24.75 | | | | |

Table 14. Cluster outputs for Miller-Nicely data

| Clusters | α | Output DC for given method | |
|---|---|---|---|
| | | single | u=.3 |
| 15-16 | 0 | 0 | 0 |
| 13,15-16 | 0 | .140 | .140 |
| 123,15-16 | 0 | .175 | .175 |
| 123,45,15-16 | 0 | .241 | .241 |
| 123,45,8-11,15-16 | 0 | .277 | .277 |
| 123,45,8-11,9-10,15-16 | .003 | .310 | .310 |
| 123,45,67,8-11-12,9-10,15-16 | .00182 | .311 | .311 |
| 123,45,67,8-11-12,9-10, 13-14,15-16 | .000834 | .328 | .328 |
| 123,45,67,8-11-12, 9-10-13-14,15-16 | .00381 | .349 | .371 |
| 123,45,67,15-16, 89-10-11-12-13-14 | .0613 | .387 | |
| 12345,67,8-11-12,15-16, 9-10-13-14 | .0144 | | .404 |
| 12345,67,15-16 89-10-11-12-13-14 | .0373 | .399 | .442 |
| 1234567,15-16, 89-10-11-12-13-14 | .0187 | .404 | .445 |
| 1234567, 89-10-11-12-13-14-15-16 | .0482 | .476 | .536 |
| 1 thru 16 | | .511 | .573 |

Table 14 (continued)

| Clusters | α | Output DC for given method u = .5 | uv=(.2,.4) |
|---|---|---|---|
| 15-16 | 0 | 0 | 0 |
| 13,15-16 | 0 | .140 | .140 |
| 123,15-16 | 0 | .175 | .175 |
| 123,45,15-16 | 0 | .241 | .241 |
| 123,45,8-11,15-16 | 0 | .277 | .277 |
| 123,45,8-11,9-10,15-16 | 0 | .297 | .297 |
| 123,45,8-11-12,9-10,15-16 | .003 | .310 | .310 |
| 123,45,67,8-11-12,9-10,15-16 | .00182 | .311 | .311 |
| 123,45,67,8-11-12,9-10 13-14,15-16 | .000834 | .328 | .328 |
| 123,45,67,8-11-12, 9-10-13-14,15-16 | .00381 | .371 | .349 |
| 123,4567,8-11-12, 9-10-13-14,15-16 | .0281 | .405 | |
| 1234567,8-11-12, 9-10-13-14,15-16 | .0377 | .445 | .404 |
| 1234567,15-16, 89-10-11-12-13-14 | .0187 | .452 | .442 |
| 1234567, 89-10-11-12-13-14-15-16 | .0482 | .545 | .509 |
| 1 thru 16 | | .586 | .567 |

Table 14 (continued)

| Clusters | α | Output DC for given method u= .7 | uv=(.2,.6) | uv=(.4,.6) | UPGMA | Complete |
|---|---|---|---|---|---|---|
| 15-16 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13,15-16 | 0 | .140 | .140 | .140 | .14 | .140 |
| 123,15-16 | 0 | .180 | .180 | .180 | .18 | .180 |
| 123,45,15-16 | 0 | .241 | .241 | .241 | .24 | .241 |
| 123,45,8-11,15-16 | 0 | .277 | .277 | .277 | .28 | .277 |
| 123,45,8-11,9-10,15-16 | 0 | .297 | .297 | .297 | .30 | .297 |
| 123,45,8-11,67,9-10,15-16 | .00112 | .311 | .311 | .311 | .31 | .311 |
| 123,45,67,8-11,9-10,13-14,15-16 | .002 | .328 | .328 | .328 | .33 | .328 |
| 123,45,67,8-11-12,9-10,13-14,15-16 | .000834 | .351 | .351 | .351 | .38 | .351 |
| 123,45,67,8-11-12,9-10-13-14,15-16 | .00381 | .371 | .371 | .371 | .38 | .417 |
| 12345,67,8-11-12,9-10-13-14,15-16 | .0144 | .443 | .436 | .438 | .43 | .452 |
| 12345,67,89-10-11-12-13-14,15-16 | .0373 | .458 | .442 | .452 | .47 | |
| 1234567,8-11-12,9-10-13-14,15-16 | .0377 | | | | | .514 |
| 1234567,89-10-11-12-13-14,15-16 | .0187 | .482 | .466 | .467 | .47 | .557 |
| 1234567,89-10-11-12-13-14-15-16 | .0482 | .559 | .553 | .553 | .54 | .593 |
| 123456789-10-11-12-13-14-15-16 | | .597 | .586 | .593 | .58 | .603 |

Table 15. Optimality measures for clustering of data from Tables 10,12,13,14.

| Table | Method | OP1 | OP2 | OP7 | OP8 | OP11 |
|---|---|---|---|---|---|---|
| 10 | single | .936 | .464 | .811 | .811 | .185 |
| | u=.3 | .939 | .323 | .811 | .811 | .141 |
| | u=.5 | .936 | .286 | .811 | .811 | .128 |
| | u=.7 | .923 | .330 | .793 | .793 | .126 |
| | complete | .912 | .620 | .718 | .718 | .231 |
| | .2,.4 | .940 | .314 | .811 | .811 | .140 |
| | .2,.6 | .928 | .301 | .793 | .793 | .114 |
| | .4,.6 | .926 | .314 | .793 | .793 | .117 |
| | UPGMA | .942 | .267 | .811 | 811 | |
| 12 | single | .754 | 1.83 | .779 | .779 | .256 |
| | u=.3 | .769 | 1.24 | .842 | .842 | .136 |
| | u=.5 | .656 | 1.17 | .805 | .805 | .134 |
| | u=.7 | .769 | 1.33 | .842 | .842 | .124 |
| | complete | .655 | 1.96 | .789 | .789 | .213 |
| | .2,.4 | .757 | 1.03 | .842 | .842 | .134 |
| | .2,.6 | .766 | 1.02 | .842 | .842 | .114 |
| | .4,.6 | .771 | 1.03 | .842 | .842 | .116 |
| | UPGMA | .771 | .958 | .842 | .842 | |
| 13 | single | .617 | 164 | .573 | .573 | .269 |
| | u=.3 | .725 | 114 | .697 | .697 | .147 |
| | u=.5 | .726 | 100 | .701 | .701 | .128 |
| | u=.7 | .726 | 107 | .717 | .717 | .149 |
| | complete | .635 | 275 | .647 | .647 | .254 |
| | .2,.4 | .732 | 115 | .701 | .701 | .148 |
| | .2,.6 | .724 | 101 | .701 | .701 | .134 |
| | .4,.6 | .729 | 97.6 | .714 | .714 | .131 |
| | UPGMA | .730 | 96.1 | .701 | .701 | |
| 14 | single | .958 | .765 | .863 | .868 | .239 |
| | u=.3 | .963 | .350 | .878 | .883 | .128 |
| | u=.5 | .960 | .349 | .863 | .868 | .110 |
| | u=.7 | .963 | .362 | .878 | .883 | .119 |
| | complete | .944 | .554 | .869 | .873 | .182 |
| | .2,.4 | .956 | .427 | .871 | .876 | .148 |
| | .2,.6 | .962 | .339 | .878 | .883 | .109 |
| | .4,.6 | .963 | .346 | .878 | .883 | .109 |
| | UPGMA | .963 | .322 | .878 | .883 | |

## REFERENCES


[1] T.S. Blyth and M.F. Janowitz, Residuation theory, London, 1972.

[2] W.W. Cooley and P.R. Lohnes, Multivariate data analysis, New York, 1971.

[3] L.A. Goodman and W.H. Kruskal, Measures of association for cross-classifications, J. Amer. Statist. Assn. 49 (1954), 732-764.

[4] J.C. Gower and C.F. Banfield, Goodness-of-fit criteria in cluster analysis and their empirical distribution. Given at the 8th International Biometric Conference, Constanta, Romania.

[5] K.J. Holzinger and H.H. Harman, Factor analysis, Chicago, 1941.

[6] L. Hubert, Some extensions of Johnson's hierarchical clustering algorithms, Psychometrika, 37(1972), 261-274.

[7] _____, Monotone invariant clustering procedures, ibid., 38(1973), 47-62.

[8] _____, Some applications of graph theory to clustering, ibid., 39(1974), 383-308.

[9] L. Hubert and F.B. Baker, Single and complete link hierarchical clustering, J. Educational Statistics, 1(1976), 87-111.

[10] M.F. Janowitz, An order theoretic model for cluster analysis, SIAM J. Appl. Math., 34(1978), 55-72.

[11] _____, Semiflat L-cluster methods, Discrete Math. 21(1978), 47-60.

[12] _____, Preservation of global order equivalence, J. Math. Psych., 20(1979), 78-88.

[13] _____, Monotone equivariant cluster methods, Siam J. Appl. Math. 37(1979), 148-165.

[14] _____, Continuous L-cluster methods, submitted for publication.

[15] _____, Similarity measures on binary attribute data, University of Massachusetts Tech. Report J7901 (1979).

[16] N. Jardine and R. Sibson, Mathematical taxonomy. New York, 1973.

[17] J.B. Kruskal, Multidimensional scaling and other methods for discovering structure, in Statistical Methods for Digital Computers, Mathematical Methods for Digital Computers, Vol. III, Enslein, Ralston and Wilf eds., New York, 1977, 296-339.

[18] G.A. Miller and P.E. Nicely, An analysis of perceptual confusions among some English consonants. J. Acoustical Society of America 27(1955), pp. 338-352.

[19] J.E. Overall, A configural analysis of psychiatric diagnostic stereotypes, Behav. Sci., 8(1963), 211-219.

[20] R.N. Shepard and P. Arabie, Additive clustering: representation of similarities as combinations of discrete overlapping properties. Psych. Reivew 86 (1979), 87-123.

[21] P.H. A. Sneath and R.R. Sokal, Numerical taxonomy, San Francisco, 1973.

[22] R.R. Sokal and F.J. Rohlf, The comparison of dendrograms by objective methods, Taxon, 11(1962), 33-40.